Swallowing sparing intensity modulated radiotherapy (SW-IMRT) in head and neck cancer

Christianen, Miranda E. M. C.; van der Schaaf, Arjen; van der Laan, Hans Paul; Verdonck-de Leeuw, Irma M.; Doornaert, Patricia; Chouvalova, Olga; Steenbakkers, Roel J. H. M.; Leemans, Charles Rene; Oosting, Sjoukje F.; van der Laan, Bernard F. A. M.

Published in:
Radiotherapy and Oncology

DOI:
10.1016/j.radonc.2015.11.009

Publication date:
2016

Radiotherapy and Oncology 118 (2016) 298–303

Contents lists available at ScienceDirect

Radiotherapy and Oncology

journal homepage: www.thegreenjournal.com

Dysphagia

Swallowing sparing intensity modulated radiotherapy (SW-IMRT) in head and neck cancer: Clinical validation according to the model-based approach ☆

Miranda E.M.C. Christianen [a], Arjen van der Schaaf [a], Hans Paul van der Laan [a], Irma M. Verdonck-de Leeuw [b], Patricia Doornaert [c], Olga Chouvalova [a], Roel J.H.M. Steenbakkers [a], Charles René Leemans [b], Sjoukje F. Oosting [d], Bernard F.A.M. van der Laan [e], Jan L.N. Roodenburg [f], Ben J. Slotman [c], Hendrik P. Bijl [a], Johannes A. Langendijk [a,*]

[a] Department of Radiation Oncology, University of Groningen, University Medical Center Groningen; [b] Department of Otorhinolaryngology-Head and Neck Surgery; [c] Department of Radiation Oncology, VU University Medical Center Amsterdam; [d] Department of Medical Oncology; [e] Department of Otorhinolaryngology-Head and Neck Surgery; and [f] Department of Oral and Maxillofacial Surgery, University of Groningen, University Medical Center Groningen, The Netherlands

ARTICLE INFO

Article history:
Received 19 July 2015
Received in revised form 6 November 2015
Accepted 18 November 2015
Available online 14 December 2015

Keywords:
Radiotherapy
Head and neck cancer
Swallowing sparing IMRT
Model based approach

ABSTRACT

Purpose: The aim of this study was to clinically validate a multivariable normal tissue complication probability (NTCP) model for grade 2–4 swallowing dysfunction at 6 months after radiotherapy or chemoradiation ($SWAL_{M6}$) in head and neck cancer patients treated with swallowing sparing intensity modulated radiotherapy (SW-IMRT) and to test if SW-IMRT resulted in a reduction of the prevalence of $SWAL_{M6}$.

Materials and methods: The primary endpoint was $SWAL_{M6}$. For all 186 patients, a standard IMRT (parotid sparing) and a SW-IMRT plan (additional constraints for swallowing organs at risk) was created. The difference in NTCP for $SWAL_{M6}$ ($\Delta NTCP_{SWALM6} = NTCP_{standard} - NTCP_{SW\text{-}IMRT}$) was calculated. Patients were treated with SW-IMRT. The external validation of the NTCP model was analyzed by comparing performance measures.

Results: The mean $\Delta NTCP_{SWALM6}$ was 4.9% (range 0.01–17.3%), with a significant lower mean predicted $NTCP_{SW\text{-}IMRT}$ of 22.6% (95% CI 20.2–24.9%), compared to $NTCP_{standard}$ of 27.5% (95% CI 24.9–29.9%) ($p < 0.001$). There was a perfect correspondence of $NTCP_{SW\text{-}IMRT}$ with the observed prevalence of $SWAL_{M6}$ (22.6%). The overall model performance, discrimination and 'goodness of fit' were good.

Conclusion: We externally validated the multivariable NTCP model for $SWAL_{M6}$ in SW-IMRT treated patients, showing reduced swallowing dysfunction by reducing the dose parameters included in this NTCP model.

Swallowing dysfunction is one of the most devastating side effects after definitive radiotherapy (RT) or chemoradiation (CHRT) for head and neck cancer (HNC) and has a major impact on health-related quality of life (HRQoL) [1–6].

Recently, we reported on the results of a large multicenter prospective cohort study in which we developed multivariable normal tissue complication probability (NTCP) models for swallowing dysfunction [7]. In that study we identified two independent risk factors for grade 2–4 swallowing dysfunction at 6 months after completion of treatment ($SWAL_{M6}$), including the mean dose to the superior pharyngeal constrictor muscle (superior PCM) and the mean dose to the supraglottic larynx, which is in line with the results reported by other investigators [8–11]. We subsequently showed that swallowing sparing intensity modulated radiotherapy (SW-IMRT) is expected to result in clinically relevant reductions in the risk of swallowing dysfunction in approximately half of the patients. SW-IMRT refers to IMRT with dose constraints for both the parotid glands as well as for the swallowing organs at risk (SWOARs), without compromising the dose to the planning target volumes (PTV) and the parotid glands [12]. However, the clinical validation of SW-IMRT remains to be determined.

Therefore, the purpose of this study was to clinically validate a previously developed multivariable NTCP model for $SWAL_{M6}$ in a cohort of patients treated with SW-IMRT and to investigate whether SW-IMRT actually resulted in a reduction of the

☆ *Funding:* Dutch Cancer Society (Grant RUG 2008-3983).

* Corresponding author at: Department of Radiation Oncology, University Medical Center Groningen, PO Box 30001, 9300 RB Groningen, The Netherlands.
E-mail address: j.a.langendijk@umcg.nl (J.A. Langendijk).

http://dx.doi.org/10.1016/j.radonc.2015.11.009
0167-8140/

prevalence of $SWAL_{M6}$ as compared to standard IMRT (parotid gland sparing IMRT).

Methods and materials

Patients

The study population of this prospective cohort study was composed of 186 consecutive patients treated from September 2010 to September 2014 at the Department of Radiation Oncology of two medical centers in The Netherlands: the VU University Medical Center (VUMC), Amsterdam or the University Medical Center Groningen (UMCG), Groningen. All patients were treated with definitive RT for squamous cell HNC originating from the oral cavity, pharynx or larynx, either alone or in combination with concomitant chemotherapy or cetuximab. All patients were subjected to a standard follow-up program that included prospective evaluation of acute and late toxicity, patient-rated symptoms and HRQoL, prior to, during and at regular intervals after treatment (NCT02435576, clinicaltrials.gov) [11,13].

Patients who previously underwent surgery, radiotherapy and/or chemotherapy and those with prior malignancies, and/or distant metastases were excluded. Patients with RTOG grade 2–4 swallowing dysfunction at baseline were also excluded in order to ensure that the observed swallowing dysfunction was induced by radiation treatment itself and not by tumor extension. Furthermore patients with residual disease or recurrence within 6 months were excluded.

Endpoints

The primary endpoint was defined as grade 2–4 swallowing dysfunction according to the RTOG/EORTC Late Radiation Morbidity Scoring Criteria, as assessed 6 months after completion of treatment ($SWAL_{M6}$), which means that patients were not able to eat solid food and were only able to eat semi-solid food, swallow liquids or were dependent on tube feeding.

Treatment

All organs at risk (OARs), including the salivary glands, and the SWOARs were delineated as previously described [14,15]. Regions of interest, IMRT planning and optimization for SW-IMRT were described previously in detail [12].

Study design and statistical analysis

The validation of SW-IMRT was performed according to the model-based approach [16]. Model-based validation means that multivariable NTCP models developed in a population with a certain radiation technique (in this case standard IMRT and three-dimensional conformal RT (3D-CRT)) are also valid in an independent subsequent population treated with another radiation technique (in this case SW-IMRT) (external validation). This implies that dose reductions (ΔDose) obtained with the new technology result in NTCP reductions (ΔNTCP) as predicted by the multivariable NTCP model and thus that the new technique indeed contributes to less toxicity. For this purpose, the following steps were made:

Step 1 ($NTCP_{standard}$): For all patients, a standard IMRT treatment plan was created, using dose constraints for the parotid glands but without dose constraints for the SWOARs. For this plan, the NTCP value for $SWAL_{M6}$ ($NTCP_{standard}$) was calculated using the equation of the prediction model for $SWAL_{M6}$ as described by Christianen et al. [7]:

$NTCP = (1 + e^{-S})^{-1}$, in which $S = -6.09 + (\text{mean dose superior PCM} * 0.057) + (\text{mean dose supraglottic larynx} * 0.037)$.

This plan was then saved and stored.

Step 2 ($NTCP_{SW\text{-}IMRT}$): This standard IMRT plan was then further optimized into a SW-IMRT plan with similar planning objectives for the parotid glands and target volumes, but with additional constraints for the SWOARs. During the planning procedures, attempts were made to reduce the dose to the SWOARs as much as possible in an iterative process, until the dose to other organs at risk started to rise (parotid glands, oral cavity or spinal cord) and/or the dose levels to the PTV's were compromised. For this plan, the NTCP value for $SWAL_{M6}$ ($NTCP_{SW\text{-}IMRT}$) was calculated with the same equation that was used in step 1.

Subsequently, $\Delta NTCP_{SWALM6}$ was calculated, defined as $NTCP_{standard} - NTCP_{SW\text{-}IMRT}$, and thus corresponding with the predicted NTCP reduction that could be obtained with SW-IMRT as compared to standard IMRT in each individual patient. Patients were divided into 2 groups, including a LOW $\Delta NTCP_{SWALM6}$ group if $\Delta NTCP_{SWALM6}$ was 0–5% and a HIGH $\Delta NTCP_{SWALM6}$ group if $\Delta NTCP_{SWALM6}$ was more than 5%. This threshold was chosen in advance on an arbitrary basis.

Step 3 (actual treatment): Patients were then actually treated with SW-IMRT and prospectively followed using exactly the same data registration program as used in the previous population treated with 3D-CRT or standard IMRT in which the multivariable model was developed [7].

Step 4 (external model validation): The external validation of the NTCP model was done by a number of performance measures of the NTCP model in the SW-IMRT cohort to that obtained in the original cohort treated with standard IMRT or 3D-CRT that was used to develop the model [7]. For this purpose, we used Monte–Carlo simulations using repeated random drawings of the outcomes according to the model NTCP to generate the expected distributions of the performance measures, based on the model and the case-mix of the cohort, and calculated single sided *p*-values (i.e., testing the null-hypothesis that the actual performance was not worse than expected). Overall model performance was described with the explained variance (using the Nagelkerkes R^2) and the scaled Brier score. The Brier score is the average squared difference between the predicted probability and the actual outcome. The scaled Brier score is a recalculated Brier score that will give a more robust comparability of the accuracy of the model. A scaled Brier score should be as close to 1 as possible (a perfect model), and is 0 for a non-informative model [17,18]. For the discrimination ability of the model, we calculated the area under the receiver operating characteristic curve (AUC), and the discrimination slope, defined as the absolute difference between the mean predicted NTCP value for patients with and without the outcome [17,18]. Finally, a Hosmer–Lemeshow goodness-of-fit test was performed to evaluate the calibration of the model. Model calibration describes to what extent the observed prevalence in a number of equally-sized subgroups corresponds with the expected values based on the average NTCP-value of each subgroup. The model's predictions fit the data at an acceptable level if the Hosmer–Lemeshow goodness-of-fit test statistic is >0.05 [19]. Since in the previous cohort [7] we only assessed model performance using the AUC, we retrospectively performed all the above mentioned model performance measures in the previous cohort as well.

Step 5 (technique validation): To evaluate the ability of SW-IMRT to reduce the prevalence of $SWAL_{M6}$, we analyzed if the observed prevalence of $SWAL_{M6}$ significantly differed from the average predicted $NTCP_{standard}$, using a non-parametric two-sided test.

Data were analyzed using SPSS Statistics for Windows, version 19.

Results

The study population of this prospective cohort study was composed of 186 patients with a mean age of 64 years. The majority of patients were males (74%). The pre-treatment characteristics of the patients are listed in Table 1.

The mean dose in all SWOARs and the contralateral parotid glands was significantly lower with the SW-IMRT plans compared to the standard IMRT plans, whereas the mean dose in the ipsilateral parotid and submandibular glands remained the same (Table 2A).

The $\Delta NTCP_{SWALM6}$ varied widely between individual patients, with a mean $\Delta NTCP_{SWALM6}$ of 4.9% (range 0.01–17.3%). Out of 186 patients, 87 (47%) were classified as having HIGH $\Delta NTCP_{SWALM6}$ while 99 patients (53%) were classified as having LOW $\Delta NTCP_{SWALM6}$ (Fig. 1). The LOW and HIGH $\Delta NTCP_{SWALM6}$ groups differed significantly with regard to a number of pretreatment variables (Table 1). The patients in the HIGH $\Delta NTCP_{SWALM6}$ group had higher T-stages, had more primary tumors originating from nasopharynx and oropharynx, and were more often treated with CHRT or conventional RT (Table 1). Moreover, the doses delivered to some of the SWOARs were significantly higher (Table 2B and C).

In the standard IMRT and 3D-CRT patient cohort used for model development [7], the overall model performance was good with a scaled Brier of 0.23 (95% CI 0.15–0.31) and an explained variance (R^2) of 0.31 (95% CI 0.21–0.41). Discrimination in terms of the AUC was 0.80 (95% CI 0.75–0.86) with a discrimination slope of 0.22 (95% CI 0.18–0.25). The Hosmer–Lemeshow "goodness of fit" had a p-value of 0.19, indicating a good agreement between expected and observed rates.

Similar results were seen in the validation of the current SW-IMRT patient cohort. The overall model performance had an actual scaled Brier of 0.13 (95% CI 0.03–0.22, with $p = 0.36$ for a single sided test with respect to the expected distribution based on the model and the case-mix, such that we cannot reject the null-hypothesis that the performance is not worse than expected) and an actual explained variance (R^2) of 0.21 (95% CI 0.08–0.33, $p = 0.34$). The discrimination ability of the model showed an actual AUC of 0.75 (95% CI 0.68–0.82, $p = 0.32$) and an actual discrimination slope with a value of 0.14 (95% CI 0.10–0.18, $p = 0.30$). The Hosmer–Lemeshow "goodness of fit" had a $p = 0.74$, indicating a good agreement between expected and observed rates.

Fig. 2 shows the calibration plots of both cohorts.

The mean predicted $NTCP_{SW\text{-}IMRT}$ in all 186 patients was 22.6% (95% CI 20.2–24.9%), which was significantly lower than the mean predicted $NTCP_{standard}$ which was 27.5% (95% CI 24.9–29.9%) ($p < 0.001$). The observed prevalence of $SWAL_{M6}$ was 22.6%, which corresponded perfectly with the mean predicted $NTCP_{SW\text{-}IMRT}$ values (Fig. 3).

In the LOW $\Delta NTCP_{SWALM6}$ group, the mean predicted $NTCP_{standard}$ and mean predicted $NTCP_{SW\text{-}IMRT}$ was 23.3% (95% CI 19.2–27.6%) and 21.3% (95% CI 17.4–25.4%), respectively ($p < 0.001$). In the LOW $\Delta NTCP_{SWALM6}$ the observed prevalence was 20.2% which corresponded with both 95%-confidence intervals of the mean predicted $NTCP_{standard}$ and $NTCP_{SW\text{-}MRT}$ (Fig. 3).

In the HIGH $\Delta NTCP_{SWALM6}$ group, the mean predicted $NTCP_{standard}$ was 32.2% (95% CI 30.2–34.3%) while the mean predicted $NTCP_{SW\text{-}IMRT}$ was significantly lower, i.e. 24.1% (95% CI 22.1–26.3%) ($p < 0.001$). The observed prevalence of $SWAL_{M6}$ in the HIGH $\Delta NTCP_{SWALM6}$ was 25.3% and was within the mean predicted $NTCP_{SW\text{-}IMRT}$ 95%-confidence interval (Fig. 3).

Discussion

This study is the first to report the clinical validation of SW-IMRT using a model-based approach. Our results show that by adding dose constraints for SWOARs during treatment planning optimization a clinically relevant ΔNTCP can be obtained in approximately 50% of the patients and that subsequent lower prevalences of $SWAL_{M6}$ are observed if the dose to the superior PCM and the supraglottic larynx can indeed be sufficiently decreased. In this regard, it should be noted that we arbitrarily defined a ΔNTCP-threshold of 5% as clinically relevant.

In the current study population, we were able to significantly reduce the dose to the superior PCM and supraglottic larynx resulting in a significant average NTCP reduction from 27.5% obtained with standard IMRT to 22.6% as obtained with SW-IMRT. The expected average $NTCP_{SW\text{-}IMRT}$ corresponded perfectly with the observed prevalence of 22.6%. These results are in line with those

Table 1
Patients characteristics.

Characteristics		All patients ($n = 186$)		LOW NTCP reduction group (ΔNTCP: 0–5%) ($n = 99$)		HIGH NTCP reduction group (ΔNTCP >5%) ($n = 87$)		P-value*
		Number	%	Number	%	Number	%	
Sex	Male	138	74	77	78	61	70	.233
	Female	48	26	22	22	26	30	
Age, years	18–65	109	59	55	56	54	62	.368
	>65	77	41	44	44	33	38	
Tumor classification	T1–T2	103	55	63	64	40	46	.016
	T3–T4	83	45	36	36	47	54	
Node classification	N0	96	52	57	58	39	45	.083
	N+	90	48	42	42	48	55	
Primary Site	Larynx	85	46	50	51	35	40	.007
	Oropharynx	64	35	27	27	37	43	
	Oral cavity	8	4	5	5	3	3	
	Hypopharynx	21	11	16	16	5	6	
	Nasopharynx	8	4	1	1	7	8	
Treatment modalities	Conventional radiotherapy	51	27	24	24	27	31	.006
	Accelerated radiotherapy	63	34	42	43	21	24	
	Chemoradiation	61	33	24	24	37	43	
	Bioradiation	11	6	9	9	2	2	

* P-value LOW NTCP reduction versus HIGH NTCP reduction, based on chi-square.

Table 2
Dose distribution parameters.

	Dose (Gy) according to actually given SW-IMRT plan		Dose (Gy) according to BACK UP standard IMRT plan		Difference between SW-IMRT and standard IMRT (Gy)	*P*-value*
	Average D_{mean}	95% CI	Average D_{mean}	95% CI		
(A) All patients						
Organ at risk						
Superior pharyngeal constrictor muscle	41.5	38.2–44.7	44.4	41.0–47.7	2.9	<.001
Middle pharyngeal constrictor muscle	46.8	43.9–49.7	50.9	48.2–53.5	4.1	<.001
Inferior pharyngeal constrictor muscle	53.7	51.4–55.7	57.4	55.4–59.2	3.7	<.001
Cricopharyngeal muscle	49.2	47.1–51.2	52.2	50.2–54.0	3.0	<.001
Esophageal inlet muscle	35.4	33.0–37.5	41.7	39.2–43.7	6.3	<.001
Supraglottic larynx	54.3	52.0–56.3	58.1	56.1–59.8	3.8	<.001
Glottic larynx	56.6	54.0–58.9	58.9	56.7–61.0	2.3	<.001
Parotid gland ipsilateral	27.8	25.4–30.4	27.9	25.6–30.4	0.1	.120
Parotid gland contralateral	20.9	19.1–22.7	21.5	19.7–23.3	0.6	<.001
Submandibular gland ipsilateral	49.6	46.0–53.1	49.8	46.2–53.2	0.2	.044
Submandibular gland contralateral	44.0	40.5–47.2	44.3	40.8–47.5	0.3	.003
(B) LOW NTCP reduction group (ΔNTCP: 0–5%)						
Structure						
Superior pharyngeal constrictor muscle	34.3	29.0–39.4	35.6	30.3–40.8	1.3	<.001
Middle pharyngeal constrictor muscle	41.0	36.2–45.8	44.6	40.1–49.1	3.6	<.001
Inferior pharyngeal constrictor muscle	54.9	51.5–57.9	57.4	54.3–60.4	2.5	<.001
Cricopharyngeal muscle	48.7	45.6–51.8	51.0	48.0–54.0	2.3	<.001
Esophageal inlet muscle	30.7	27.1–34.1	36.0	32.4–39.2	5.3	<.001
Supraglottic larynx	54.2	50.4–57.1	56.8	53.2–59.7	2.6	<.001
Glottic larynx	59.1	55.6–62.4	60.1	56.7–63.1	1.0	<.001
Parotid gland ipsilateral	21.9	18.4–25.3	21.8	18.4–25.2	0.1	.591
Parotid gland contralateral	16.5	13.7–19.2	16.8	14.0–19.5	0.3	.006
Submandibular gland ipsilateral	39.9	34.1–45.5	40.3	34.5–45.7	0.4	.002
Submandibular gland contralateral	34.7	29.2–39.6	35.0	29.5–39.9	0.3	.008
(C) HIGH NTCP reduction group (ΔNTCP >5%)						
Structure						
Superior pharyngeal constrictor muscle	49.7	46.9–52.5	54.4	52.3–56.6	4.7	<.001
Middle pharyngeal constrictor muscle	53.6	51.7–55.8	58.2	56.7–59.8	4.6	<.001
Inferior pharyngeal constrictor muscle	52.4	49.8–55.5	57.4	55.4–59.6	5.0	<.001
Cricopharyngeal muscle	49.8	47.5–52.4	53.5	51.6–55.5	3.7	<.001
Esophageal inlet muscle	40.7	38.4–43.0	48.2	46.6–49.9	7.5	<.001
Supraglottic larynx	54.4	51.9–56.8	59.4	57.6–61.2	5.0	<.001
Glottic larynx	53.7	50.8–57.2	57.6	55.3–60.3	3.9	<.001
Parotid gland ipsilateral	34.5	32.0–37.1	34.9	32.5–37.3	0.4	.002
Parotid gland contralateral	26.0	24.3–27.8	27.0	25.2–28.8	1.0	<.001
Submandibular gland ipsilateral	60.9	59.1–62.5	60.8	59.1–62.4	0.1	.628
Submandibular gland contralateral	54.8	52.6–56.6	54.9	52.9–56.7	0.1	.165

† *P*-value LOW NTCP reduction versus HIGH NTCP reduction, based on independent sample *t*-test.
* *P*-value actually given dose SW-IMRT versus BACK UP standard IMRT, based on paired sample *t*-test.

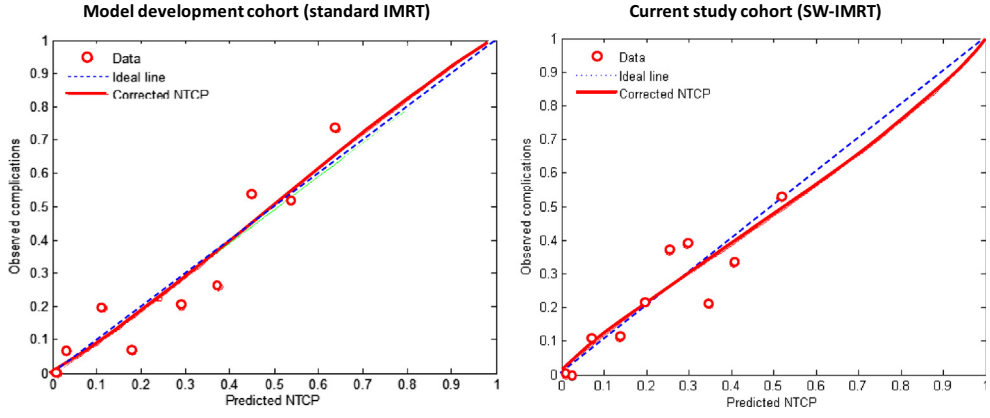

Fig. 1. NTCP comparison plot. The plot shows the NTCP values for SW-IMRT as a function of the NTCP values for standard IMRT. All dots are below the black dashed line indicating that the NTCP values for SW-IMRT are generally lower than those obtained with standard IMRT. The blue squares indicate patients with ⩾ΔNTCP >5% and the red dots indicate patients with ΔNTCP 0–5%. *Abbreviations:* IMRT = intensity modulated radiotherapy, SW-IMRT = swallowing sparing IMRT, NTCP = normal tissue complication probability.

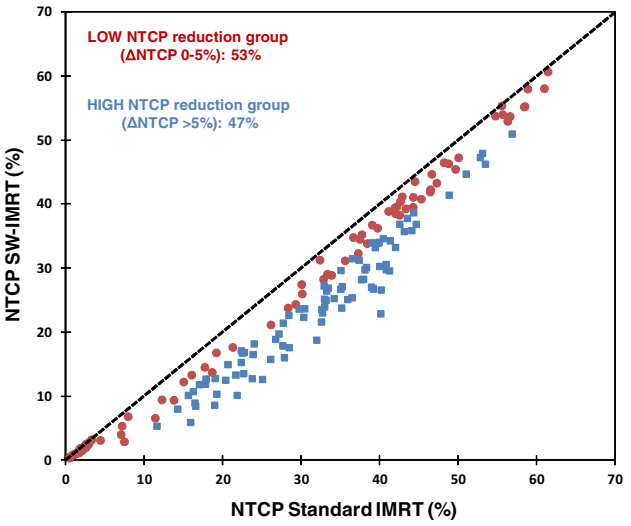

Fig. 2. Calibration plots for the previous cohort [7] treated with standard IMRT versus that for the current cohort treated with SW-IMRT. *Abbreviations:* IMRT = intensity modulated radiotherapy, SW-IMRT = swallowing sparing IMRT, NTCP = normal tissue complication probability.

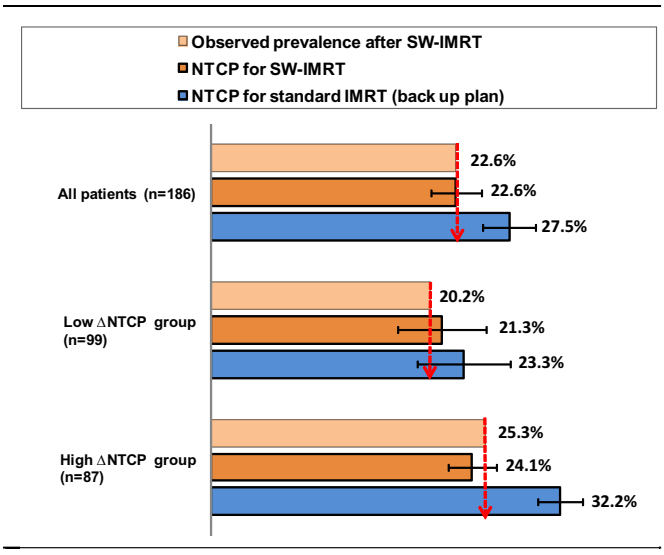

Fig. 3. Observed prevalence after SW-IMRT (dark blue) compared to predicted NTCP values for SW-IMRT (light blue) and to predicted NTCP values for standard IMRT based on the back up standard IMRT plans (orange). The observed prevalences in all patients and in those in the HIGH ΔNTCP groups corresponded significantly better with the average NTCP values for SW-IMRT than with those for standard IMRT (red arrow). *Abbreviations:* IMRT = intensity modulated radiotherapy, SW-IMRT = swallowing sparing IMRT, NTCP = normal tissue complication probability.

reported by Feng et al. who observed low rates of dysphagia at 1 year after chemoradiation after IMRT with sparing of the pharyngeal constrictors and glottic regions [20,21].

It should be noted that in this study, we only included patients with primary tumors originating from the oral cavity, nasopharynx, oropharynx, hypopharynx and larynx, while in our previous cohort [7] also other primary tumor sites were included. When confining the analysis to similar tumor sites as used in the current study, the prevalence of $SWAL_{M6}$ in patients treated with standard IMRT in our previous study was 27.9%, which corresponded nicely with the mean predicted $NTCP_{standard}$ found in the current study (27.5%), indicating that the expected prevalence of $SWAL_{M6}$ would have been similar to that observed in the previous cohort when treated with standard IMRT.

The NTCP for $SWAL_{M6}$ with standard IMRT and the $\Delta NTCP_{SWALM6}$ varied widely between individual patients. We found differences between the pretreatment variables of the patients in the LOW (0–5%) and HIGH (>5%) $\Delta NTCP_{SWALM6}$ groups. The HIGH $\Delta NTCP_{SWALM6}$ group, consisted of patients with more advanced T-stages, and with more nasopharyngeal and oropharyngeal cancers, which corresponds with the higher doses administered to the most important SWOARs, i.e. the superior PCM and supraglottic larynx. Consequently, patients in the HIGH $\Delta NTCP_{SWALM6}$ group more often received CHRT or conventional RT.

As shown in a previous report on the implementation of SW-IMRT [22], ΔNTCP obtained with SW-IMRT depend on uni- versus bilateral neck RT, tumor location and the amount of overlap SWOAR-PTV. Also in the current study, we experienced several reasons why $\Delta NTCP_{SWALM6}$ remained low in some of the patients. In the first place, the SWOARs sometimes partly or even completely overlapped with the PTV, and therefore little to no reduction in the SWOARs could be obtained without compromising the dose to the PTV. Secondly, lowering the dose to the SWOARs in some cases led to an increase in the dose to other OARs which was not allowed according to predefined criteria. Finally, for a small subgroup of patients the initial NTCP was already low and could not be reduced any further.

Recently, Vainshtein et al. reported on favorable long-term patient-reported outcome of swallowing complaints among HPV-positive oropharyngeal cancer patients treated with chemoradiation with IMRT with dose constraints for both salivary glands and a number of swallowing structures, which further support the validity of IMRT aimed at the reduction of dysphagia [23]. The relationship between dose–volume parameters and swallowing dysfunction after RT or CHRT, and on the potential benefit of sparing SWOARs with IMRT, but the clinical relevance of these reductions remained to be determined [2,21–26]. Some of the other investigators who previously reported on the potential dosimetric benefits of SW-IMRT, accepted reduced coverage of the (elective) PTV (i.e. by using split field IMRT) [22,26,27–30], which makes it difficult to compare these results with those from the current study.

In the current study we used the previously described model-based approach to validate SW-IMRT. The model-based approach is a stepwise methodology, that has been developed to effectively select patients that are expected to benefit most from new radiation delivery techniques aiming at the reduction of radiation-induced side effects, such as proton therapy [16]. However, this method is also applicable for the development and validation of other radiation delivery techniques, such as SW-IMRT.

At present, randomized trials are still considered gold standard in evidence-based medicine. This is certainly true for new interventions aiming at improving treatment efficacy in terms of local control and survival. However, evidence based medicine is not restricted to randomized trials and meta-analyses. It involves tracking down the best external evidence with which to answer our clinical questions [31]. For new radiation techniques that are only aiming at reducing side effects without changing the strategy with regard to tumor control (i.e. target volumes and fractionation), the model-based approach can be considered as a good alternative for an RCT.

The model-based approach consists of two phases (phase α): a phase aiming at the development of a new radiation technique (3 steps), and a consecutive phase (phase β) including a prospective observational cohort study aiming at the clinical validation of the new radiation delivery technique (step 4). For swallowing dysfunction and SW-IMRT, phase α, including multivariable NTCP model development followed by in silico planning comparisons have been

reported in our previously published papers [7,12]. In the current paper we reported on phase β. The multivariable NTCP model for $SWAL_{M6}$ showed a significant relationship between the dose to the superior PCM and supraglottic larynx and $SWAL_{M6}$. However, such relationship does not strictly guarantee a causal relationship between these two dose parameters and $SWAL_{M6}$. Therefore, an essential step in assessing the generalizability and causality of the multivariable NTCP model for $SWAL_{M6}$ was external validation in an independent patient cohort treated with a radiation technique in which the relevant dose parameters were further optimized [32]. The current study showed that the multivariable NTCP model for $SWAL_{M6}$ still performed well in this independent patient cohort with a modified IMRT technique and that the predictions based on the actual treatment technique (SW-IMRT) corresponded very well with the observed prevalence."

In conclusion, we externally validated the multivariable NTCP model for $SWAL_{M6}$ in a subsequent independent cohort of patients treated with SW-IMRT and showed that by reducing the dose parameters included in this NTCP model, the risk of swallowing dysfunction can be reduced.

Conflict of interest

The department of Radiation Oncology of the UMCG has research agreements with Philips, RaySearch and Mirada.

Acknowledgements

This study was financially supported by a grant from the Dutch Cancer Society (Grant RUG 2008-3983).

References

[1] Goguen LA, Posner MR, Norris CM, et al. Dysphagia after sequential chemoradiation therapy for advanced head and neck cancer. Otolaryngol Head Neck Surg 2006;134:916–22.
[2] Jensen K, Lambertsen K, Grau C. Late swallowing dysfunction and dysphagia after radiotherapy for pharynx cancer: frequency, intensity and correlation with dose and volume parameters. Radiother Oncol 2007;85:74–82.
[3] Langendijk JA, Doornaert PA, Verdonck-de Leeuw IM, Leemans CR, Aaronson NK, Slotman BJ. Impact of late treatment-related toxicity on quality of life among patients with head and neck cancer treated with radiotherapy. J Clin Oncol 2008;26:3770–6.
[4] Nguyen NP, Frank C, Moltz CC, et al. Impact of dysphagia on quality of life after treatment of head-and-neck cancer. Int J Radiat Oncol Biol Phys 2005;61:772–8.
[5] Ramaekers BLT, Joore MA, Grutters JPC, et al. The impact of late treatment-toxicity on generic health-related quality of life in head and neck cancer patients after radiotherapy. Oral Oncol 2011;47:768–74.
[6] Vergeer MR, Doornaert PA, Rietveld DHF, Leemans CR, Slotman BJ, Langendijk JA. Intensity-modulated radiotherapy reduces radiation-induced morbidity and improves health-related quality of life: results of a nonrandomized prospective study using a standardized follow-up program. Int J Radiat Oncol Biol Phys 2009;74:1–8.
[7] Christianen MEMC, Schilstra C, Beetz I, et al. Predictive modelling for swallowing dysfunction after primary (chemo)radiation: results of a prospective observational study. Radiother Oncol 2012;105:107–14.
[8] Wopken K, Bijl HP, Van-der-Schaaf A, Van-der-Laan HP, Chouvalova O, Steenbakkers RJ, et al. Development of a multivariable normal tissue complication probability (NTCP) model for tube feeding dependence after curative radiotherapy/chemo-radiotherapy in head and neck cancer. Radiother Oncol 2014;113:95–101.
[9] Mortensen HR, Jensen K, Aksglæde K, Behrens M, Grau C. Late dysphagia after IMRT for head and neck cancer and correlation with dose-volume parameters. Radiother Oncol 2013;107:288–94.
[10] Sanguineti G, Rao N, Gunn B, Ricchetti F, Fiorino C. Predictors of PEG dependence after IMRT±chemotherapy for oropharyngeal cancer. Radiother Oncol 2013 Jun;107:300–4.
[11] De Ruyck K, Duprez F, Werbrouck J, Sabbe N, Sofie de L, Boterberg T, et al. A predictive model for dysphagia following IMRT for head and neck cancer: introduction of the EMLasso technique. Radiother Oncol 2013 Jun;107:295–9.
[12] Van der Laan HP, Christianen MEMC, Bijl HP, Schilstra C, Langendijk JA. The potential benefit of swallowing sparing intensity modulated radiotherapy to reduce swallowing dysfunction: an in silico planning comparative study. Radiother Oncol 2012;103:76–81.
[13] Beetz I, Schilstra C, Burlage FR, et al. Development of NTCP models for head and neck cancer patients treated with three-dimensional conformal radiotherapy for xerostomia and sticky saliva: the role of dosimetric and clinical factors. Radiother Oncol 2012;105:86–93.
[14] Christianen MEMC, Langendijk JA, Westerlaan HE, van de Water TA, Bijl HP. Delineation of organs at risk involved in swallowing for radiotherapy treatment planning. Radiother Oncol 2011;101:394–402.
[15] Van de Water TA, Bijl HP, Westerlaan HE, Langendijk JA. Delineation guidelines for organs at risk involved in radiation-induced salivary dysfunction and xerostomia. Radiother Oncol 2009;93:545–52.
[16] Langendijk JA, Lambin P, De Ruysscher D, Widder J, Bos M, Verheij M. Selection of patients for radiotherapy with protons aiming at reduction of side effects: the model-based approach. Radiother Oncol 2013;107:267–73.
[17] Gerds TA, Cai T, Schumacher M. The performance of risk prediction models. Biom J 2008;50:457–79.
[18] Steyerberg EW, Vickers AJ, Cook NR, et al. Assessing the performance of prediction models: a framework for some traditional and novel measures. Epidemiology 2010;21:128–38.
[19] Hosmer DW, Hosmer T, Le Cessie S, Lemeshow S. A comparison of goodness-of-fit tests for the logistic regression model. Stat Med 1997;16:965–80.
[20] Feng FY, Kim HM, Lyden TH, Haxer MJ, Worden FP, Feng M, et al. Intensity-modulated chemoradiotherapy aiming to reduce dysphagia in patients with oropharyngeal cancer: clinical and functional results. J Clin Oncol 2010 Jun 1;28:2732–8.
[21] Eisbruch A, Kim HM, Feng FY, Lyden TH, Haxer MJ, Feng M, et al. Chemo-IMRT of oropharyngeal cancer aiming to reduce dysphagia: swallowing organs late complication probabilities and dosimetric correlates. Int J Radiat Oncol Biol Phys 2011 Nov 1;81:e93–9.
[22] Van der Laan HP, Gawryszuk A, Christianen MEMC, et al. Swallowing-sparing intensity-modulated radiotherapy for head and neck cancer patients: treatment planning optimization and clinical introduction. Radiother Oncol 2013;107:282–7.
[23] Vainshtein JM, Moon DH, Feng FY, Chepeha DB, Eisbruch A, Stenmark MH. Long-term quality of life after swallowing and salivary-sparing chemo-intensity modulated radiation therapy in survivors of human papillomavirus-related oropharyngeal cancer. Int J Radiat Oncol Biol Phys 2015 Apr 1;91:925–33.
[24] Caglar HB, Tishler RB, Othus M, et al. Dose to larynx predicts for swallowing complications after intensity-modulated radiotherapy. Int J Radiat Oncol Biol Phys 2008;72:1110–8.
[25] Eisbruch A, Schwartz M, Rasch C, et al. Dysphagia and aspiration after chemoradiotherapy for head-and-neck cancer: which anatomic structures are affected and can they be spared by IMRT? Int J Radiat Oncol Biol Phys 2004;60:1425–39.
[26] Fua TF, Corry J, Milner AD, Cramb J, Walsham SF, Peters LJ. Intensity-modulated radiotherapy for nasopharyngeal carcinoma: clinical correlation of dose to the pharyngo-esophageal axis and dysphagia. Int J Radiat Oncol Biol Phys 2007;67:976–81.
[27] Teguh DN, Levendag PC, Noever I, et al. Treatment techniques and site considerations regarding dysphagia-related quality of life in cancer of the oropharynx and nasopharynx. Int J Radiat Oncol Biol Phys 2008;72:1119–27.
[28] Caudell JJ, Burnett OL, Schaner PE, Bonner JA, Duan J. Comparison of methods to reduce dose to swallowing-related structures in head and neck cancer. Int J Radiat Oncol Biol Phys 2010;77:462–7.
[29] Schwartz DL, Hutcheson K, Barringer D, et al. Candidate dosimetric predictors of long-term swallowing dysfunction after oropharyngeal intensity-modulated radiotherapy. Int J Radiat Oncol Biol Phys 2010;78:1356–65.
[30] Webster GJ, Rowbottom CG, Ho KF, Slevin NJ, Mackay RI. Evaluation of larynx-sparing techniques with IMRT when treating the head and neck. Int J Radiat Oncol Biol Phys 2008;72:617–22.
[31] Sackett D, Rosenberg W, Muir Gray J, Haynes RB, Richardson WS. Evidence based medicine: what it is and what it isn't. Br Med J 1996;312:71–2.
[32] Van der Schaaf A, Langendijk JA, Fiorino C, Rancati T. Embracing phenomenological approaches to Normal Tissue Complication Probability modeling: a question of method. Int J Radiat Oncol 2015;91:468–71.